文章编号:1672-3961(2024)05-0093-08 DOI:10.6040/j.issn.1672-3961.0.2023.150

基于盖根堡多项式最佳平方近似的谱图网络

林振宇,邵蓥侠*
(北京邮电大学计算机学院, 北京 100876)

摘要:针对现有谱图神经网络模型在学习图节点特征矩阵信号频率分布方面存在的不足,采用盖根堡正交基改进,提出一种泛化能力强、适合真实世界数据的谱图神经网络模型,有效提高节点分类任务精度。分析不同真实世界数据集中图节点特征矩阵的信号频率分布,使用盖根堡正交基学习谱图滤波函数,提高模型的泛化能力。理论分析表明,该模型能够以最佳平方误差有效学习闭区间上的任意连续谱滤波函数。在 13 个数据集上进行试验的结果显示,基于盖根堡正交基的谱图神经网络模型在 8 个数据集上的性能均超越目前的先进模型,验证了模型的有效性。可扩展性试验表明,该模型适用于大规模图数据。

关键词:盖根堡正交基;谱图神经网络;图节点特征矩阵;信号频率分布;滤波函数
中图分类号:TP391 **文献标志码**:A
引用格式:林振宇,邵蓥侠. 基于盖根堡多项式最佳平方近似的谱图网络[J]. 山东大学学报(工学版),2024,54(5):93-100.
LIN Zhenyu, SHAO Yingxia. Spectral graph networks based on best square approximation with Gegenbauer polynomials[J]. Journal of Shandong University (Engineering Science), 2024, 54(5):93-100.

Spectral graph networks based on best square approximation with Gegenbauer polynomials

LIN Zhenyu, SHAO Yingxia*
(School of Computer Science, Beijing University of Post and Telecommunication, Beijing 100876, China)

Abstract: To address the limitations of existing spectral graph neural network models in learning the frequency distribution of signals in graph node feature matrices, a Gegenbauer-based spectral graph neural network model with strong generalization ability was proposed, suitable for real-world data, which effectively improved node classification accuracy. The signal frequency distribution in graph node feature matrices from various real-world datasets was analyzed, and a method using the Gegenbauer orthogonal basis to learn spectral graph filtering functions was proposed, enhancing the model's generalization ability. Theoretical analysis demonstrated that the model was capable of effectively learning arbitrary continuous spectral filtering functions on closed intervals with the best square error. Experiments conducted on 13 datasets showed that the performance of the Gegenbauer-based spectral graph neural network model surpassed advanced models on 8 out of 13 datasets, which confirmed the model's effectiveness. Scalability experiments indicated that the model was applicable to large-scale graph data.

Keywords: Gegenbauer orthogonal basis; spectral graph neural network; graph node feature matrix; signal frequency distribution; filtering function

0 引言

谱图神经网络是一种基于图信号处理和深度学习的图神经网络模型[1],主要思想是将傅里叶变换应用到图数据上,提取每个节点的局部特征,在谱域上进行卷积运算。早期的谱图神经网络使用各阶特征值矩阵的线性组合近似卷积核参数[2]。传统谱图模型需要进行图拉普拉斯矩阵的对角化分解,所需内存规模和计算量会随节点数量快速增长,在训练过程中,每次前向传播都需要复杂计算,并且训练参数正比于图的规模,导致传统谱图神经

收稿日期:2023-05-29
基金项目:国家自然科学基金资助项目(62272054, 62192784)
第一作者简介:林振宇(2000—),男,浙江台州人,硕士研究生,主要研究方向为图神经网络。E-mail:linzy@bupt.edu.cn
*通信作者简介:邵蓥侠(1988—),男,浙江宁波人,教授,博士生导师,博士,主要研究方向为大规模图处理。E-mail:shaoyx@bupt.edu.cn

网络难以训练或应用到大规模图数据(拥有100 000及以上节点或边的图数据)上。为克服上述困难,已有谱图神经网络的相关研究尝试使用一组多项式基线性加权组合近似需要学习的复杂参数[3-12]。但现有研究存在2个方面的问题。一方面,现有模型使用的多项式基泛化能力不足,缺乏选取多项式基的相关理论,例如:切比雪夫网络(Chebyshev network, ChebNet)基于切比雪夫多项式,理论泛化能力较好,能以最佳一致损失逼近任何滤波器[3],但在真实世界数据集上的效果却劣于图卷积网络(graph convolutional network, GCN)[13],目前研究仍不能很好地解释这一现象;克雷网络(Cayley network, CayleyNet)[4]、自回归移动平均(auto-regressive moving average, ARMA)滤波器[5]、泛化页排序图神经网络(generalized pagerank graph neural network, GPRGNN)[6]、伯恩斯坦网络(Bernstein network, BernNet)[7]等基于不同类型的非正交多项式近似任意谱滤波函数,尽管节点分类性能较好,但泛化能力均弱于基于正交多项式的先进模型(如雅可比卷积网络(Jacobi convolutional network, JacobiConv)[12]、ChebNet)。因此,关于谱图神经网络设计过程中如何选取一组泛化能力更强的多项式基,目前仍缺乏明确的理论指导。另一方面,现有谱图模型在设计时没有考虑节点特征矩阵的频率分布。图最重要的2个信息就是图拓扑结构和节点特征矩阵。目前,基于谱图理论的工作将拉普拉斯矩阵变换到频域,却没有研究过节点特征信号在频域上的分布情况,在设计模型时也忽略了节点特征与频率分布的关联,解决这一问题有助于形成更优的谱图神经网络。

为了解决上述2个问题,本研究从函数近似理论出发,指出基于多项式的模型以最小平方误差近似谱滤波函数时,多项式基需要满足正交条件,为后续谱图神经网络的设计提供有效的理论工具;通过分析正交多项式的权重函数和真实世界数据集中节点特征频率分布的关联,提出盖根堡正交多项式谱图神经网络模型,更适应真实世界数据集的特征频率分布;试验结果表明本研究所提模型在选用的13个真实世界数据集中,有8个的表现均超过目前先进的谱图神经网络模型,证明理论分析和模型设计的有效性。扩展性试验表明,本研究提出的模型能够有效应用在大规模图上。

1 相关工作

谱图神经网络在处理同配图和异配图数据时都十分有效,在大规模数据集上有较好的可扩展性,因此具有广泛应用和重要研究价值[14-23]。

1.1 谱图理论与傅里叶变换

谱图神经网络利用卷积定理,通过傅里叶变换将参与卷积的2个函数转换到频域进行相乘,使用傅里叶逆变换将结果转换回空域。

将N节点的无向连通图表示为$G=(V,E)$,其中V为图中节点集合,E为图中边集合,滤波函数为g_θ,节点特征矩阵为$\boldsymbol{X}$,$\boldsymbol{X}\in\mathbf{R}^{N\times d}$,拉普拉斯矩阵$\boldsymbol{L}=\boldsymbol{I}-\boldsymbol{A}$,其中$\boldsymbol{I}$为单位矩阵,$\boldsymbol{I}\in\mathbf{R}^{N\times N}$,$\boldsymbol{A}$为邻接矩阵,$\boldsymbol{A}\in\mathbf{R}^{N\times N}$,分解后的特征值矩阵$\boldsymbol{\Lambda}\in\mathbf{R}^{N\times N}$,特征向量矩阵$\boldsymbol{U}\in\mathbf{R}^{N\times N}$,$\boldsymbol{L}=\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}$,$\boldsymbol{\Lambda}$中位置$(i,i)$处的元素为$\lambda_i$,$\lambda_i\in[0,2]$,$\forall i\in[1,n]$。特征矩阵$\boldsymbol{X}$的傅里叶变换和傅里叶逆变换分别表示为:

$$\hat{\boldsymbol{X}}=\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X},\tag{1}$$

$$\boldsymbol{X}=\boldsymbol{U}\hat{\boldsymbol{X}},\tag{2}$$

式中$\hat{\boldsymbol{X}}$为谱域上的节点特征矩阵。根据卷积定理,一个图神经网络模型可以从谱域角度表达为:

$$g_\theta(\boldsymbol{L})*\boldsymbol{X}=\boldsymbol{U}(g_\theta(\boldsymbol{\Lambda})(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}))=\boldsymbol{U}g_\theta(\boldsymbol{\Lambda})\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X},\tag{3}$$

式中,$*$为卷积运算。训练谱图神经网络就是学习定义域为$[0,2]$的g_θ在n个特征点上的$g_\theta(\lambda_i)$,因此,可以转化为训练含有n个参数的函数θ_i,即:

$$\boldsymbol{U}\begin{pmatrix}g_\theta(\lambda_1)&\cdots&0\\ \vdots&&\vdots\\ 0&\cdots&g_\theta(\lambda_n)\end{pmatrix}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}=\boldsymbol{U}\begin{pmatrix}\theta_1&\cdots&0\\ \vdots&&\vdots\\ 0&\cdots&\theta_n\end{pmatrix}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}。\tag{4}$$

谱图神经网络虽然具有较好的可解释性,却存在2个问题:$\boldsymbol{L}$的特征分解需要巨大的计算量和内存开销;需要学习的参数正比于图中节点的规模,因此难以扩展到大规模图上。

1.2 基于多项式的谱图神经网络

为了克服上述问题,研究者们利用标准分解后特征向量矩阵的特殊性质$\boldsymbol{U}\boldsymbol{U}^{\mathrm{T}}=\boldsymbol{I}$[24],引入高阶多项式近似卷积核函数,具体原理式为:

$$g(\boldsymbol{L})=\boldsymbol{U}\left(\sum_{k=0}^{K}\theta_k\boldsymbol{\Lambda}^k\right)\boldsymbol{U}^{\mathrm{T}}=\sum_{k=0}^{K}(\theta_k\boldsymbol{U}\boldsymbol{\Lambda}^k\boldsymbol{U}^{\mathrm{T}})=\sum_{k=0}^{K}\theta_k(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}})\cdots(\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}})=\sum_{k=0}^{K}\theta_k\boldsymbol{L}^k,\tag{5}$$

式中K为谱图神经网络的阶数。该方法使模型参数量从$O(N)$降低到$O(K)$。进一步使用k阶多项式函数$P_k(\boldsymbol{L})$代替$\boldsymbol{L}^k$,以求更好地近似$g(\boldsymbol{L})$函

数。最终,通用的基于多项式近似的谱图神经网络(spectral graph neural network, SGNN)模型表示为:

$$\boldsymbol{Z}=\left(\sum_{k=0}^{K}\theta_k P_k(\boldsymbol{L})\right)\boldsymbol{X}。\tag{6}$$

近年来,基于多项式近似谱图滤波函数的模型都可看作式(6)的特殊情况,主要可以分为以下两类。

第一类是使用预定义或特殊形式的权重系数组合多阶多项式的模型[3]。该类模型的提出并非从谱域视角出发,但仍可将其看作谱图神经网络的一种。为了让空域图神经网络模型有效扩展到大规模图数据,文献[8]提出简化图卷积(simplifying graph convolutional, SGC)网络,可以看作使用 K 阶多项式 $(1-x)^K$ 近似谱图滤波器,权重系数为 1;文献[9]使用可调参的 PageRank 系数构造近似个性化网页排名网络(approximate personalized pagerank of neural prediction, APPNP);文献[10]使用固定系数分别学习图上的低频和高频信号,提出低频/高频-图神经网络(graph neural network-low frequency/graph neural network-high frequency, GNN-LF/GNN-HF),按照特定比例平衡低频谱图滤波函数和高频谱图滤波函数。

第二类是使用可学习的权重系数组合多项式的模型,最早是 ChebNet[2];CayleyNet 用 Cayley 多项式的实数域部分[4];ARMA 将多个有理式组合成增强谱滤波器,用于拟合不平滑图信号[5];为了提升图神经网络在异配图上的效果,GPRGNN 模型使用幂函数基作为多项式基,假设对应的权重系数可正可负[6];文献[3]指出先前的工作使用预定义或无约束的多项式权重线性组合多个多项式滤波器,导致滤波器过分简化,因此,文献[7]使用 Bernstein 多项式作为基,约定权重系数非负。

上述两类模型中多项式泛化能力不足,没有考虑图节点特征在频域上的信号分布,限制了模型的表达性能。

1.3　已有研究的不足

文献[12]分析了选用不同正交基对谱图神经网络模型收敛速率的影响,提出基于正交雅可比多项式的权重可学习的 JacobiConv,在常见数据集上均达到先进模型的效果,但该模型同 1.2 节中提到的模型一样,没有考虑节点特征矩阵的频率分布,因此模型能力还有进一步提高的空间。另外,目前已有研究均未给出选用多项式设计谱图神经网络时的理论指导。

2　正交多项式基的最佳平方误差近似

2.1　多项式基的选取

定理 1　魏尔斯特拉斯逼近定理[25]　假设 f 为一个定义在 $[a,b]$ 上的连续函数,对 $\forall\varepsilon>0$,存在多项式 p,使 $\|f(x)-p(x)\|_\infty<\varepsilon$。

由定理 1 可知,任意多项式都能以特定的精度逼近给定区间上的连续函数。但在实际谱图神经网络模型中,选取一种性质良好的多项式,可以让模型更准确高效地学习谱滤波函数,而性质不好的多项式会引起一系列问题,例如采用幂函数基求滤波函数的最佳平方逼近时,训练过程中模型精度剧烈变化,近似函数舍入误差很大,可以通过病态系数 C_p 进行分析:

$$C_p=\frac{\left|\dfrac{\Delta f(x)}{f(x)}\right|}{\left|\dfrac{\Delta x}{x}\right|},\tag{7}$$

式中,$\Delta f(x)$ 为值域变化量,Δx 为定义域变化量。当 f 代表的幂函数阶次较大时,C_p 较大。

定理 2　最佳平方逼近定理[24,26-27]　假设谱滤波函数为 $[0,2]$ 上的连续函数,一组正交多项式基为 $\{P_0,P_1,\cdots,P_K\}$,任意不超过 K 阶的 $[0,2]$ 上的函数都能由这组基的线性组合表示,则 g 在子集 $\{P_0,P_1,\cdots,P_K\}$ 上的最佳平方逼近

$$f^*=\underset{f}{\operatorname{argmin}}\|g-f\|_2^2=\underset{f\in P}{\operatorname{argmin}}\left\|g-\sum_{k=0}^{K}\alpha_k P_k\right\|_2^2,\tag{8}$$

式中 α_k 为权重系数。根据定理 2,正交多项式能最佳平方逼近任意谱滤波函数,因此具有很强的泛化能力。选取正交多项式具有 2 点优势:一组正交多项式任意两项之间的互信息为 0,任意连续函数可以唯一表达为正交多项式的线性组合,不存在交叉项,因此使用 K 阶多项式拟合谱滤波函数只需要确定 K 个系数,大大降低模型的参数量和训练难度;任意一组正交多项式各项之间存在递推关系,可以利用递推公式简化计算。因此正交多项式最常用于近似任意连续函数,选取正交多项式成为设计谱图神经的关键要素。

2.2　正交基权重函数在图上的意义

为了选取合适的正交多项式,从数学定义出发,引出权重函数概念,解释权重函数在真实世界数据上的含义。通过图像直观展示不同权重函数在拟合特征频率分布时的能力。正交的定义与权

重函数 $\omega(x)$ 有关，$\omega(x)$ 代表 2 个函数计算乘积时在不同维度的权重，是一个非负的连续函数。

若 $f(x),g(x)\in[a,b]$ 满足

$$\langle f(x),g(x)\rangle=\int_{-1}^{1}\omega(x)f(x)g(x)\mathrm{d}x, \quad (9)$$

则称 $\langle f(x),g(x)\rangle$ 为 $f(x)$ 与 $g(x)$ 在 $[a,b]$ 上的带权内积。如果式(9)等于 0，则称 $f(x)$ 与 $g(x)$ 关于权重函数 $\omega(x)$ 正交。

若函数族 $\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)\}$ 满足

$$\langle\varphi_i(x),\varphi_j(x)\rangle=\int_a^b\omega(x)\varphi_i(x)\varphi_j(x)\mathrm{d}x=\begin{cases}0, & i\neq j\\ A_i, & i=j\end{cases}, \quad (10)$$

则称 $\{\varphi_k(x)\}$ 为 $[-1,1]$ 上带权 $\omega(x)$ 的正交函数族，其中 A_i 为正交基 φ_i 的模长，$A_i>0$。每类正交基有对应的权重函数，Chebyshev 正交基的权重函数为 $1/\sqrt{1-x^2}$。任意一个权重函数通过选定合适的初始函数族，经过施密特正交化，即可得到一组对应的正交基。因此，确定一组合适的正交基需要结合数据集特征选择合适的权重函数。

$\omega(x)$ 和真实图数据集有着紧密的联系，在图上 $\omega(\lambda)$ 为信号频率分布密度，其中 λ 为频率，离散情况下可以表示为 $[\lambda_{i-1},\lambda_i]$ 之间的信号频率分布函数的变化率，即：

$$\omega(\lambda)=\frac{\Delta F(\lambda)}{\Delta\lambda}=\frac{\sum_{\lambda_i\leqslant\lambda}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X})_i^2-\sum_{\lambda_{i-1}\leqslant\lambda}(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X})_i^2}{\lambda_i-\lambda_{i-1}}, \quad (11)$$

式中：$\Delta F(\lambda)$ 为信号频率分布函数的变化；$\Delta\lambda$ 为频率的变化；λ_1、λ_2、$\cdots$、λ_n 为 n 个频率，$\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$；$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}$ 为特征矩阵在对应频率上的幅值。由于在大图上难以通过矩阵对角化过程得到对应的 $\boldsymbol{U}^{\mathrm{T}}$，所以本研究仅在 cora、pubmed、citeseer、chameleon、squirrel、film 等小规模图中绘制特征频率分布。

在图节点特征信号中，某些频率分量上的幅度远远大于其他幅度，为了便于展示，在绘图时本研究将这些点过滤掉，对信号进行均值滤波，使其变得更加平滑，真实世界数据集上的特征信号频率分布如图 1 所示。图 1 直观地表明 Chebyshev 权重函数难以拟合真实世界图上的特征频率分布。节点特征信号分布中存在的大幅值点也表明，在将特征矩阵 $\boldsymbol{X}$ 输入图网络之前，可以先随机丢弃一些噪声，让正交多项式的权重函数更好地拟合特征频率分布。

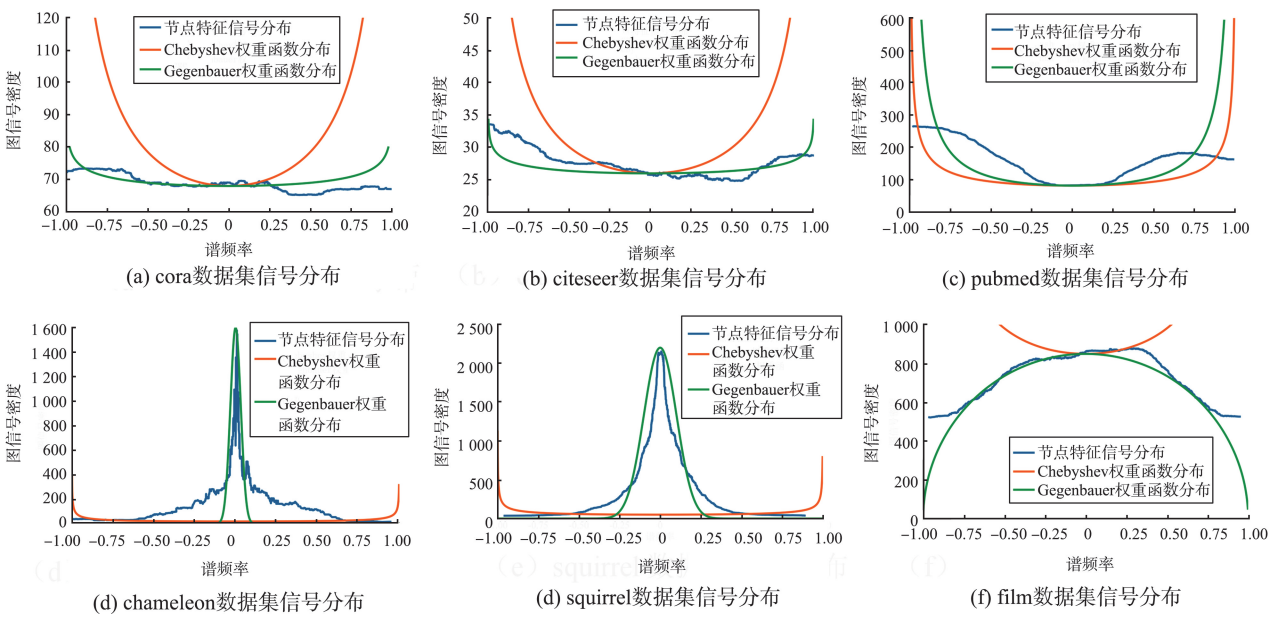


图 1 真实世界数据集上的特征信号频率分布

Fig.1 Graph signal frequency distribution on real-world datasets

2.3 基于盖根堡多项式的谱图模型

盖根堡网络（Gegenbauer network，GegenNet）是式(6)框架下的一种情况，将真实数据集节点特征在频域上的分布结合到谱图神经网络设计中，提出的基于盖根堡正交多项式的谱图神经网络具有最优平方误差的近似能力，比已有工作拥有更加强大的泛化能力。

2.2 节中提到 Chebyshev 多项式对应的权重函数为 $1/\sqrt{1-x^2}$，但根据图 1 中的分析结果，真实世界的节点特征频率分布并不一定满足这个形式。因此，本研究通过向权重函数中引入一个可学习参数 a，提出 Gegenbauer 多项式，其权重函数 $\omega(x)=(1-x^2)^{a-\frac{1}{2}}$，由图 1 可以看出，相比 Chebyshev，Gegenbauer 权重函数

的拟合能力显著增强。

Gegenbauer 多项式 $\{C_0^a, C_1^a, \cdots, C_{n+1}^a\}$ 满足的递推关系可以表示为:

$$C_0^a(\boldsymbol{L}) = \boldsymbol{I}, \tag{12}$$

$$C_1^a(\boldsymbol{L}) = 2a\boldsymbol{L}, \tag{13}$$

$$C_{n+1}^a(\boldsymbol{L}) = \frac{1}{n+1}[2\theta_n \boldsymbol{L} C_n^a(\boldsymbol{L}) - \theta_{n-1} C_{n-1}^a(\boldsymbol{L})], \tag{14}$$

式中:θ_n、θ_{n-1}为递推公式中的多项式系数,$\theta_n = n+a$,$\theta_{n-1} = n+2a-1$。引入 Gegenbauer 多项式基后,提出的 GegenNet 模型可以表示为:

$$\boldsymbol{Z}_{\mathrm{G}} = \sum_{k=0}^{K} (\alpha_k C_k^a(\boldsymbol{L}))\boldsymbol{X}。 \tag{15}$$

根据 2.2 节可知,节点特征矩阵中在某些频率点上的权重远远超过其他部分,这些点为噪声点,为便于分析,本研究将这些点忽略。所以 GegenNet 模型在将特征输入谱图网络前,对 $\boldsymbol{X}$ 的部分噪声进行丢弃,利用递推公式加快训练和推理过程。GegenNet模型算法如下。

算法 1　GegenNet 模型算法

输入:K、a、n、$\boldsymbol{L}$、$\boldsymbol{X}$。

输出:$\boldsymbol{Z}_{\mathrm{G}} = \sum_{k=0}^{K} (\alpha_k C_k^a(\boldsymbol{L}))\boldsymbol{X}$。

(1) 初始化 $\boldsymbol{Z}_{\mathrm{G}}$;

(2) for $k=0$ to K do;

(3) $\alpha_k = 1$;

(4) if $k==0$ or $k==1$ then;

(5)　$C_0^a(\boldsymbol{L}) = \boldsymbol{I}$, $C_1^a(\boldsymbol{L}) = 2a\boldsymbol{L}$;

(6)　end if;

(7)　else then;

(8)　$C_k^a(\boldsymbol{L}) = (2\theta_{k-1}\boldsymbol{L}C_{k-1}^a(\boldsymbol{L}) - \theta_{k-2}C_{k-2}^a(\boldsymbol{L}))/k$;

(9)　end else;

(10)　$\boldsymbol{Z}_{\mathrm{G}} = \boldsymbol{Z}_{\mathrm{G}} + \alpha_k C_k^a(\boldsymbol{L})$;

(11) end for;

(12)　$\boldsymbol{Z}_{\mathrm{G}} = \boldsymbol{Z}_{\mathrm{G}}\boldsymbol{X}$;

(13) return $\boldsymbol{Z}_{\mathrm{G}}$。

3　试验分析

3.1　泛化能力评估试验

本试验通过不同模型在图像数据集上拟合不同类型滤波器时的误差损失评价模型之间的泛化能力。试验设置参考文献[8]。

从 MATLAB 图像处理工具包中取出 50 张 100 像素×100 像素的图片构成数据集,每张图像看成节点出度为 4 的网格图(边缘节点出度为 3),像素强度转换为 10 000 维的向量。为了评估不同谱图神经网络的泛化能力,选用 5 类常用的信号滤波函数(低通滤波器为$\mathrm{e}^{-10\lambda^2}$,高通滤波器为 $1-\mathrm{e}^{-10\lambda^2}$,宽带滤波器为$\mathrm{e}^{-10(\lambda-1)^2}$,窄带滤波器为$\mathrm{e}^{-10(\lambda-1)^2}$,梳型滤波器为$|\sin\ \pi\lambda|$)作用在所有像素的结果上作为标签值,使用不同的谱图神经网络模型拟合这些滤波器函数,采用平均均方误差损失 L_{MSE} 作为损失函数:

$$L_{\mathrm{MSE}} = \sum_{n=1}^{N} \sqrt{(y_{\mathrm{pred}} - y_{\mathrm{label}})^2}, \tag{16}$$

式中,y_{pred}为 GegenNet 输出的预测值,y_{label}为其他基线输出的标签值。通过比较损失,反映不同谱图神经网络模型的泛化能力。试验中,采用一些先进的谱图神经网络模型作为基线。

试验结果如表 1 所示,其中最优结果加粗标记,次优结果用下划线标记。由表 1 可知:除窄带滤波器外,在拟合各个类型的滤波器时,GegenNet 模型的平均均方误差损失都达到最低,其中拟合低通滤波器和高通滤波器时,损失均为10^{-4}量级;拟合宽带滤波器时,损失较其他模型最小;拟合梳型滤波器时,损失略优于 JacobiConv 模型;拟合窄带滤波器时,损失也接近 JacobiConv 模型。该试验证明 GegenNet模型的泛化性能超越了基线模型。

表 1　在 50 张图像上的平均均方误差损失

Table 1　The sum of mean squared error loss over 50 images

模型	平均均方误差损失				
	低通滤波器	高通滤波器	宽带滤波器	窄带滤波器	梳型滤波器
GCN	3.479 9	67.663 5	25.875 5	21.074 7	50.512 0
GPRGNN	0.416 9	0.094 3	3.512 1	3.791 7	4.654 9
ARMA	1.847 8	1.863 2	7.692 2	8.273 2	15.121 4
ChebNet	0.822 0	0.786 7	2.272 2	2.529 6	4.073 5
BernNet	0.031 4	0.011 3	0.041 1	0.931 3	0.998 2
JacobiConv	0.000 3	0.001 1	0.021 3	**0.015 6**	0.293 3
GegenNet	**0.000 2**	**0.000 2**	**0.015 6**	0.018 6	**0.286 7**

3.2 小规模图数据集上的节点分类试验

本试验采用常见的小规模数据集,既有同配图数据集(Cora、Citeseer、Pubmed),也有异配图数据集(Chameleon、Squirrel、Film、Texas、Cornell),大部分节点标签与邻居标签不同。小规模图数据集的统计信息如表 2 所示。本研究采用文献[13]中的划分比例,训练集、验证集、测试集比例分别为 60%、20%、20%。试验结果采用 20 轮试验的平均结果,每轮的随机种子采用随机初始化。

表 2 小规模图数据集统计信息

Table 2 Statistics of the small scalegraph datasets

数据集	节点数	边数	特征维度	类别数	同质度
Cora	2 708	5 429	1 433	7	0.81
Citeseer	3 327	4 732	3 703	6	0.74
Pubmed	19 717	44 338	500	3	0.80
Chameleon	2 277	36 101	2 325	5	0.23
Squirrel	5 201	217 073	2 089	5	0.22
Film	7 600	33 544	931	5	0.22
Texas	183	309	1 703	5	0.11
Cornell	183	295	1 703	5	0.23

节点分类准确率

$$a_{\mathrm{cc}}=\frac{\sum_{n=1}^{N} I(y_{\mathrm{pred}}=y_{\mathrm{label}})}{N}, \tag{17}$$

式中,I 为指示函数,当 $y_{\mathrm{pred}}=y_{\mathrm{label}}$ 时,指示函数值为 1,否则为 0。试验结果如表 3 所示,其中最优结果加粗标记,次优结果用下划线标记。由表 3 可以看出:GegenNet 在 Cora、Citeseer、Pubmed、Chameleon、Film 数据集上的节点分类准确率达到最佳;在 2 个异配图数据集 Film、Chameleon 上,节点分类准确率相比 BernNet 分别提升近 6 百分点、7 百分点;在 Citeseer 数据集上,节点分类准确率相比 JaocbiConv 提升 3 百分点。该试验结果证明了正交多项式基模型的有效性。GegenNet 模型在这些数据集上之所以有效,可以从图 1 获得解释,图 1 中的绿色曲线代表 GegenNet 权重函数在取不同参数时拟合图数据集信号频率分布的结果,通过选取不同的参数,GegenNet均能够较好地拟合这些数据集的特征。

表 3 在小规模图数据集上节点分类准确率

Table 3 Accuracy of node classification on the small scale graph dataset

模型	节点分类准确率/%							
	Cora	Citeseer	Pubmed	Chameleon	Squirrel	Film	Texas	Cornell
GPRGNN	86.70±1.03	75.12±1.98	87.38±0.63	67.28±1.09	36.47±1.38	50.15±1.92	84.59±4.37	**92.97±1.68**
GCN	81.14±1.01	79.86±0.67	86.73±0.27	59.61±2.21	46.78±0.87	33.23±1.16	77.38±3.28	65.90±4.43
ChebNet	74.91±0.52	67.69±0.64	65.91±1.71	37.15±1.49	26.58±1.92	26.55±0.46	36.35±8.90	28.78±4.85
BernNet	88.52±0.95	80.09±0.79	88.48±0.41	68.29±1.58	**41.79±1.01**	51.35±0.73	**93.12±0.65**	92.13±1.64
JacobiConv	88.98±0.46	80.78±0.79	89.62±0.41	74.20±1.03	41.17±0.64	57.38±1.25	93.44±2.13	92.95±2.46
GegenNet	**89.41±0.08**	**83.15±0.34**	**89.65±0.04**	**74.94±0.77**	40.41±0.19	**59.80±0.91**	90.16±1.64	88.29±5.32

由表 3 可知,对于 Texas 和 Cornell 数据集,GegenNet模型也达到了较好的效果,节点分类准确率优于 ChebNet、GCN 等模型。该试验结果可以从信号频率分布和权重函数的角度出发进行解释。Cornell 数据集上的图信号密度分布和权重函数拟合结果如图 2 所示,通过分析 Cornell 数据集上的图信号密度分布可以看出,其分布很难被单峰的权重函数拟合,这一结果可以用于证明权重函数理论在实际应用中的有效性。同理,在分析 Texas 数据集上的试验效果时,也可以使用相同的方法,模型的节点分类准确率较低是由于权重函数和图信号频率分布不匹配。另外,由表 3 可以看出,由于 Conrell 和 Texas 数据集的规模极小(仅有 100 多个节点),基于正交多项式的谱图模型泛化能力难以完全体现,GegenNet 和 JacobiConv 模型的效果不如基于非正交多项式构造的 BernNet 模型。

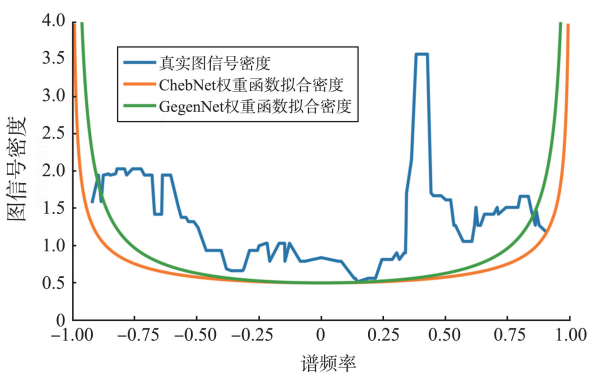


图 2 Cornell 数据集上的图信号密度分布和权重函数的拟合结果

Fig.2 The density distribution of graph signals and the fitting results of weight functions on the Cornell dataset

3.3 大规模图数据集上的节点分类试验

在大规模图数据集上进行节点分类任务,其中大规模图数据集包括大型引文网络 Ogbn-arxiv 及其他 4 个大规模异配图数据集(Penn94、Genius、Twitch-gamer、Pokec)[28]。大规模图数据集的统计

信息如表 4 所示。在 Ogbn-arxiv 数据集上,采用文献[12]中使用的 10 个随机种子划分数据集,另外 4 个异配图数据集采用给定的比例划分。

表 4　大规模图数据集统计信息
Table 4　Statistics of the large scale graph datasets

数据集	节点数	边数	特征维度	类别数	同质度
Ogbn-arxiv	169 343	1 166 243	40	40	0.42
Penn94	41 554	1 362 229	5	2	0.47
Genius	421 961	984 979	12	2	0.62
Twitch-gamer	168 114	6 797 557	7	2	0.09
Pokec	1 632 803	30 522 564	65	2	0

试验结果如表 5 所示,其中最优结果加粗标记,次优结果用下划线标记。由表 5 可以看出:在 5 个大规模图数据集上,GegenNet 模型的节点分类准确率相比基线模型有提升或相差不大;在 Penn94 数据集上,GegenNet 模型超过 BernNet 模型近 1 百分点;在 Ogbn-arxiv 和 Twitch-gamer 数据集上也超过先进模型的性能;虽然在 Genius 和 Pokec 数据集上,GegenNet 模型没有达到最优性能,但相比 GCN 和 ChebNet 模型均有较大提升,与图 1 的拟合结果相符,再次从试验角度证明了理论和试验分析的有效性。试验结果证明了 GegenNet 模型的可扩展性。

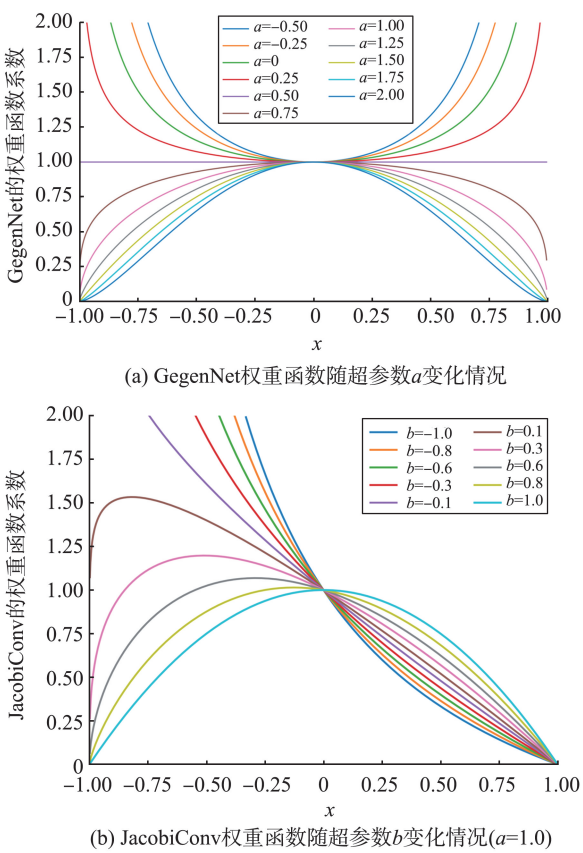
表 5　在大规模图数据集上节点分类准确率
Table 5　Accuracy of node classification on the large scale graph dataset

模型	节点分类准确率/%				
	Ogbn-arxiv	Penn94	Genius	Twitch-gamer	Pokec
GPRGNN	71.57±0.22	83.22±0.51	90.09±0.31	62.71±0.35	80.46±0.34
GCN	71.32±0.31	82.35±0.18	86.29±1.22	62.33±0.15	75.53±0.21
ChebNet	71.09±0.17	81.13±0.45	87.51±0.62	62.55±0.28	77.68±0.14
BernNet	71.96±0.27	83.26±0.29	**90.47±0.33**	64.27±0.31	**81.67±0.17**
GegenNet	**72.15±0.24**	**84.18±0.15**	88.07±0.22	**64.64±0.35**	78.51±0.18

3.4　GegenNet 与 JacobiConv 权重函数的对比

JacobiConv 模型是目前先进模型之一,在 3.2 节中,GegenNet 模型在 Cora、Citeseer、Pubmed、Chameleon、Squirrel、Film 共 6 个数据集上都超过或接近该模型。考虑 GegenNet 和 JacobiConv 都基于正交多项式基且具有较好的泛化能力,为了比较两者的不同,本研究给出 2 个模型权重函数图像,结合图 1 中数据集节点特征的信号频率分布图像和权重函数图像进一步分析 2 种模型的不同。

GegenNet 权重函数能够拟合的频率分布如图 3(a)所示,JacobiConv 权重函数如图 3(b)所示。虽然 JacobiConv 权重函数 $(1-x)^a(1+x)^b$ 包含函数族的形状比 GegenNet 对应的权重函数 $(1-x)^{a-\frac{1}{2}}(1+x)^{a-\frac{1}{2}}$ 更为丰富,但两者的权重函数都是单峰函数,因此,在 Cornell、Texas 数据集(双峰或多峰分布)上,两者都没有较好的拟合能力,从 3.2 节的试验结果中也可以证明这一点。在图 1 包含的小规模真实数据集上,特征信号频率分布形状比较符合图 3(a),因此在训练时,GegenNet 模型相比 JacobiConv,权重函数更容易收敛到特定数据集的特征频率分布,在这些数据集上达到更高的精度,与表 3 的试验结果相符。



(a) GegenNet权重函数随超参数a变化情况

(b) JacobiConv权重函数随超参数b变化情况(a=1.0)

图 3　GegenNet 与 JacobiConv 对应权重函数随参数变化

Fig.3　The weight functions corresponding to GegeNet and JacobiConv as parameters vary

4 结论

本研究解决了使用多项式基近似任意谱图滤波函数的谱图神经网络中存在的2个主要问题,提出基于最佳平方近似误差的正交多项式谱图模型。通过分析正交多项式基的权重函数和节点特征在频域的分布情况,探究ChebNet在真实世界图上效果较差的根本原因,即权重函数和真实世界数据集信号密度函数不匹配。使用泛化能力更强的权重函数对应的正交多项式提高模型拟合任意谱图滤波函数的能力。目前的研究主要局限于处理具有单一峰值的密度函数数据。然而,本研究的理论分析为未来工作提供了指导,允许研究者根据真实数据集的特定频率分布特性和正交多项式的权重函数评估模型的泛化能力。此外,理论分析的深入还可能揭示新的模型,进一步推动该领域的研究进展。

基于多项式的谱图神经网络研究还不是很充分,未来研究应该包括如何根据数据集隐含的权重函数自适应选择一组正交基构造新的谱图神经网络。探究谱图神经网络中多个滤波器的权重系数应满足的约束条件也是未来的主要研究方向。

参考文献:

[1] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[EB/OL]. (2014-05-21)[2023-03-26]. https://doi.org/10.48550/arXiv.1312.6203.

[2] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]//Proceedings of the 30th Conference on Neural Information Processing Systems. Barcelona, Spain: MIT, 2016: 3844-3852.

[3] HE M, WEI Z, WEN J R. Convolutional neural networks on graphs with Chebyshev approximation, revisited[C]// Advances in Neural Information Processing Systems. Los Angeles, USA: MIT, 2022: 7264-7276.

[4] LEVIE R, MONTI F, BRESSON X, et al. CayleyNets: graph convolutional neural networks with complex rational spectral filters[J]. IEEE Transactions on Signal Processing, 2018, 67(1): 97-109.

[5] BIANCHI F M, GRATTAROLA D, LIVI L, et al. Graph neural networks with convolutional ARMA filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(7): 3496-3507.

[6] CHIEN E, PENG J, LI P, et al. Adaptive universal generalized pagerank graph neural network[EB/OL]. (2021-10-26)[2023-03-26]. https://doi.org/10.48550/arXiv.2006.07988.

[7] HE M, WEI Z, XU H. BernNet: learning arbitrary graph spectral filters via Bernstein approximation[C]// Advances in Neural Information Processing Systems. [S.l.]: MIT, 2021: 14239-14251.

[8] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks[C]//International Conference on Machine Learning. Long Bench, USA: ACM, 2019: 6861-6871.

[9] KLICPERA J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: graph neural networks meet personalized pagerank[C]//International Conference on Learning Representations. New Orleans, USA: ICLR, 2019: 2667-2682.

[10] ZHU M, WANG X, SHI C, et al. Interpreting and unifying graph neural networks with an optimization framework[C]//Proceedings of the Web Conference 2021. Ljubljana, Slovenia: ACM, 2021: 1215-1226.

[11] CHEN Z, CHEN F, ZHANG L, et al. Bridging the gap between spatial and spectral domains: a unified framework for graph neural networks[J]. ACM Computing Surveys, 2023, 50(5): 1-42.

[12] WANG X, ZHANG M. How powerful are spectral graph neural networks[C]// International Conference on Machine Learning. Baltimore, USA: ACM, 2022: 23341-23362.

[13] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]//International Conference on Learning Representations. Toulon, France: ICLR, 2017: 48550.

[14] TANG J, LI J, GAO Z, et al. Rethinking graph neural networks for anomaly detection[C]//International Conference on Machine Learning. Baltimore, Maryland: ACM, 2022: 21076-21089.

[15] GAO Y, WANG X, HE X, et al. Alleviating structural distribution shift in graph anomaly detection[C]//Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. Singapore: ACM, 2023: 357-365.

[16] JIANG D, WU Z, HSIEH C Y, et al. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models[J]. Journal of Cheminformatics, 2021, 13(1): 1-23.

[17] TIAN Z, BAI T, ZHANG Z, et al. Directed acyclic graph factorization machines for CTR prediction via knowledge distillation[C]//Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. Singapore: ACM, 2023: 715-723.

[18] LIU M, GAO H, JI S. Towards deeper graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. [S.l.]: ACM, 2020: 338-348.

(下转第110页)